

Title	＜研究ノート＞因果推論の理論と分析手法
Author(s)	大久保, 将貴
Citation	年報人間科学. 37 p. 35-p. 49
Issue Date	2016-03-31
oaire:version	VoR
URL	https://doi.org/10.18910/54584
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

〈研究ノート〉

因果推論の理論と分析手法

大久保 将貴

要旨

本稿の目的は、因果推論の理論と分析手法を紹介し、さらに、因果推論の方法をめぐる議論を検討することである。因果推論とは、文字通り因果関係を識別することを目的としており、近年の社会科学では様々な洗練された分析手法が登場している。本稿では、因果推論の代表的な手法である、ランダム化比較実験、回帰分析、操作変数法、回帰不連続デザイン(RDD)、差の差法(DiD)について概観する。さらに、そうした手法をめぐる批判や議論について検討する。

キーワード

因果推論、ランダム化比較実験、操作変数法、RDD、DiD

1. 因果関係の識別

本稿では、因果関係を識別するための理論と方法を概観し、さらに、今日の因果推論をめぐる批判や議論を紹介し検討する。経済学や政治学では因果推論の多用が目立つ一方で、社会学における因果推論の浸透は極端に遅れている。なぜ遅れているのかについては、(1) 因果推論の方法を知らながらも社会学の目的と馴染まないため採用しない、(2) 単に因果推論を知らないから採用していない、という2つの考え方があるだろう。なぜ社会学において因果推論が浸透していないのかという問いは興味深く、社会的に検討の余地があるが、本稿ではこの点に立ち入らず、具体的な因果推論の方法を提示することに力点をおく。

1.1 相関あるいは因果

相関と因果が異なることは、社会調査や計量分析の基礎知識であり、講義でも導入部分で紹介されることが多いだろう。しかしながら社会学ではいざ分析となると、アウトカム Y を適切なコントロール変数 X を統制しながら興味のある説明変数 D に回帰した結果得られる推定量をさも因果効果のように解釈している分析例が少なくない。こうした方法は、因果関係の識別にこだわる学派からは相関関係の分析として解釈される¹⁾。相関ではなく因果を特定したい場合には、従来の回帰分析ではなく、特別な手続きが必要となるのである。まずは、Angrist and Pischke(2009) にしたがって、今日の計量分析において因果関係を特定するための4課題を先に明示する。

その1：興味のある因果関係とは何か

記述的な分析は重要な役割を担うが、一般に、因果推論において重視されるのは原因が結果に及ぼす効果を明らかにする分析である。したがって原因 D と結果 Y を明確にする必要がある。因果関係に基づいて規定される因果状態 (causal states)²⁾ を適切に定義することは、因果推論の出発点である。仮に、社会保険適用の効果に関心があり、社会保険適用がランダムに割り当てられている状況を想像しよう。社会保険適用を受けた人たちは処置群 (treatment group: $D = 1$)、受けていない人たちは対照群 (control group: $D = 0$) となり、平均因果効果 (average treatment effect: ATE) は母集団の対象者全員が「処置群に割り当てられた際の結果」と「対照群に割り当てられた際の結果」の差の平均である。以上から興味のある因果関係とは何かを問われれば、分析者は必ず以下の点を明示しなければならない。すなわち、a) D は Y を引き起こすのか、b) D が Y を引き起こすとしたらその効果の大きさはどの程度か、c) $D = d'$ を与えられた人が仮に $D = d''$ を与えられていたら Y はどの程度違うか、の3点である。

その2：興味のある因果関係を特定するのに理想的な実験は何か

実験を念頭に調査観察データを分析するべきだという主張は、社会学でも古くからなされてきた (Stouffer 1950)。理想的な実験を想像することは、興味のある変数をいかに調査観察データにおいて操作化するか、また分析における母集団は何かを熟慮する契機となる。Angrist and Pischke (2009) が指摘するように、「(仮想的に研究倫理や予算約を無視した) 何でもありの世界で問いに答えることのできる理想的な実験を思いつくことができないのであれば、現実の予算制約と調査観察データを用いて因果効果を導ける可能性はかなり低い」のである³⁾。

その3：因果関係を特定するための識別戦略 (identification strategy) は何か

識別戦略とは、調査観察データを用いたうえで、理想的な実験に近似する方法を意味する (Angrist and Krueger 1999)。現代では、非常に洗練された識別戦略が多く存在し、その代表例を後述の2章から5章で扱う。

その4：因果関係を特定するための統計的推論の方法は何か

統計的推論とは分析対象の母集団、扱う標本、標本誤差を求める際におく仮定を記述することだ。統計的推論はデータがクラスター化またはグループ化されているときに複雑となる。統計的推論の問題は決して刺激的ではないし、しばしば非常に技術的であるが、よく考え抜かれ、発想が面白いプロジェクトでさえ、その成功は統計的推論の詳細にかかっているのである (Rubin 1991 ; Angrist and Pischke 2009)。盲目的にホワイトの頑健標準誤差や標準誤差をクラスター化すれば良いという話ではないのだ。

因果関係を推論するには、上記の課題をまずクリアしなければならない。

1.2 ランダム化比較実験

一部の論者を除けば、興味のある変数をランダムに割り当てること、すなわちランダム化比較実験は因

果関係を識別するための最強のツールである。無論、調査観察データの扱いが主である社会学においてランダム割当が可能となるケースは稀であるが、ここでは、なぜ変数をランダムに割り当てることが因果関係の識別において最も強力なツールとなるのかを確認する。社会保険が適用されている人が皆無の世界を仮想し、社会計画者によって社会保険の適用がランダムに割り当てられる場合を例にし、個人 i についての社会保険適用($D_i = 1$) が健康(Y_i) に与える影響を考えよう。社会保険適用を受けた人たち（処置群）の条件付き期待値は $E[Y_i|D_i = 1]$ となり、社会保険適用を受けていない人たち（対照群）の条件付き期待値は $E[Y_i|D_i = 0]$ となる。ここで、社会保険の適用はランダム割り当てと仮定しているので $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$ である。したがって、社会保険適用の因果効果は以下で定式される。

$$\begin{aligned}
 & E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
 &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
 &= E[Y_{0i} + \tau|D_i = 1] - E[Y_{0i}|D_i = 0] \\
 &= \tau + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
 &= \tau
 \end{aligned} \tag{1}$$

ランダム割り当てでは $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$ が成立するため、処置群と対照群におけるアウトカムの平均値の差が因果効果となる一方で、ランダム割り当てを伴わない調査観察研究では $E[Y_{0i}|D_i = 1] \neq E[Y_{0i}|D_i = 0]$ となり、これがいわゆるセレクションバイアスである⁴⁾。「社会科学で実験はできない」という言葉はかつて聞かれたが、いまや幻である。近年の社会科学における実験研究の飛躍は目を見張るものがある。なお、ランダム割り当てを伴う実験研究でも、処置群と対照群の2群で共編変量等に有意な差がないかを確認するバランスチェックは必須であることを付言しておく。

2. 回帰

ランダム化比較実験ができないからといって、我々は因果関係の特定を諦めるわけにはいかない。実験研究が困難であった社会科学において、因果関係の特定に最も多用されたのが回帰(regression)であろう。しかしながら、回帰分析から因果効果を得るためには、いくつかの強い仮定が必要となるが、そうした仮定が満たされていることはまずない。以下では、回帰を用いた因果分析を概観しよう。

2.1 欠落変数バイアス

因果推論において重視される推定量は一致性である。一致性とは、推定量 $\hat{\beta}$ がサンプルサイズが増加するにつれてパラメータ β に確率収束することを意味する。以下のモデルを考えよう。

$$Health_i = \alpha_1 + \beta_1 PInsu_i + e_i \tag{2}$$

このモデルは、健康($Health$) を民間保険加入($PInsu$) に回帰したものであるが、説明変数 $Insu$ と誤差項 u_i に相関がある場合には推定量 β_1 は一致推定量とならない⁵⁾。誤差項 e_i は説明変数で説明されないその他のすべての要因を含むため、多くのコントロール変数を投入したところで誤差項と説明変数が相関をもたないモデルを設定するのは不可能に近い。1960～80年代のアメリカ計量社会学の牽引者たちはパス解析

が一致性をもたないことを意識しており、回帰係数が因果効果のように解釈されていることに危機感を覚え、特にダンカンが回帰分析の濫用を嘆いていた(Duncan 1984; Xie 2007)。しかしながら、欠落変数バイアスがある場合、モデルを拡張することでそのバイアスを測定することができる。

$$Health_i = \alpha_2 + \beta_2 Plnsu_i + \beta_3 Gene_i + e_i \quad (3)$$

遺伝子($Gene_i$)は健康に影響を与えているかもしれないし、先祖ががんを罹患しているから民間がん保険に加入しているかもしれない。この場合、モデル(2)では欠落変数が存在するがそのバイアスは以下となる。

$$\text{欠落変数バイアス} = Plnsu \text{ と } Gene \text{ の関連} \times \beta_3 \quad (4)$$

$Plnsu$ と $Gene$ の関連は、

$$Gene_i = \pi_0 + \pi_1 Plnsu_i + u_i \quad (5)$$

であることから式(4)に代入し以下を得る。

$$\text{欠落変数バイアス} = \pi_1 \times \beta_3 = \beta_1 - \beta_2 \quad (6)$$

欠落変数バイアスの公式(6)を用いれば、分析におけるモデル拡張で欠落変数について一定の考察ができるが、説明変数と誤差項の相関がなくなることはないため、因果効果の特定とまではいかない。

2.2 不適切なコントロール変数

欠落変数バイアスを確認したところで、限られた変数と分析手法しか使えない場合には、適切なコントロール変数を投入し、限定的な因果推論をするしかない。説明変数を追加することで、回帰係数の推定量を因果効果として解釈できる可能性は高まるが、だからといってアウトカムに影響を与えうる変数すべてを投入するのは、因果推論において誤りである。不適切なコントロール変数とは、着目する説明変数の被説明変数となりうる変数を意味する。具体的には、着目する説明変数と関連があると思われ、かつ時系列が後に決定している変数が不適切なコントロール変となる。同様に、因果推論における適切なコントロール変数とは、着目する説明変数が決まる時点で既に確定している変数となる。例えば、Angrist and Pischke(2009)は大卒が賃金に与える影響をみる際に、職種をコントロールするのは不適切なコントロール変数の典型だとしている。賃金を Y_i 、大卒ダミーを D_i 、職種であるホワイトカラーダミーを W_i とすると、潜在的なアウトカムを用いて以下となる。

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i} \quad (7)$$

$$W_i = D_i W_{1i} + (1 - D_i) W_{0i} \quad (8)$$

W_i をコントロールした場合の大卒が賃金に与える効果は、

$$\begin{aligned} & E[Y_i | W_i = 1, D_i = 1] - E[Y_i | W_i = 1, D_i = 0] \\ &= E[Y_{1i} | W_{1i} = 1, D_i = 1] - E[Y_{0i} | W_{0i} = 1, D_i = 0] \end{aligned} \quad (9)$$

となる。 D_i がランダムに割り当てなら $\{Y_{1i}, W_{1i}, Y_{0i}, W_{0i}\}$ と D_i は独立なので、

$$\begin{aligned}
& E[Y_{1i} | W_{1i} = 1, D_i = 1] - E[Y_{0i} | W_{0i} = 1, D_i = 0] \\
& = E[Y_{1i} | W_{1i} = 1] - E[Y_{0i} | W_{0i} = 1]
\end{aligned} \tag{10}$$

これは違うグループの比較をしていることと同義である。

$$\begin{aligned}
& E[Y_{1i} | W_{1i} = 1, D_i = 1] - E[Y_{0i} | W_{0i} = 1, D_i = 0] \\
& = E[Y_{1i} - Y_{0i} | W_{1i} = 1] + \{E[Y_{0i} | W_{1i} = 1] - E[Y_{0i} | W_{1i} = 0]\}
\end{aligned} \tag{11}$$

上式の意味は以下の通りである。ホワイトカラーの大卒者とそれ以外の人の賃金の差は、大卒でホワイトカラーの人の賃金に与える因果効果 $E[Y_{1i} - Y_{0i} | W_{1i} = 1]$ と、大卒という経験はホワイトカラーになる人々の構成を変化させるという事実を反映したセレクションバイアス $\{E[Y_{0i} | W_{1i} = 1] - E[Y_{0i} | W_{1i} = 0]\}$ の項との和になっている。社会学では、上記のような不適切なコントロール変数が回帰モデルに投入されていることがしばしばある。それは特定の変数の因果効果に関心がなく、モデルフィットに関心があるからだと言われればその通りであるが、因果推論の作法では不適切なコントロール変数が明確に位置付けられることを確認した。データ等の制限で現代的な因果推論の方法が使えず、シンプルな回帰分析しかできない場合、それでも回帰式に因果関係の解釈を望むならば、先の賃金モデルに職種（ホワイトカラー）を含めてはならないのである。

3. 操作変数

これまでは以下で概観するような現代的な因果推論の方法が使えない場合に、なんとか因果関係に踏み込む方法をみた。しかしながら、実際にはOLS 推定をもって因果効果と解釈する査読論文は皆無だろう。因果効果の識別に多用されている代表的手法が操作変数法である。以下のモデルを考える。

$$Health_i = \beta_0 + \beta_1 Insu_i + e_i \tag{12}$$

これは健康 (*Health*) が社会保険加入 (*Insu*) によって決まっているというモデルである。先に触れたように、 $Insu_i$ と e_i に相関がある場合に β_1 は一致推定量とならず、因果効果として解釈できない。操作変数法とは、ここに少し手を加えて一致推定量を得るための方法である。仮に社会保険加入の権限がくじ引きによってランダムに決定されるとしよう。くじに当選しても加入しないことも選択できるため、加入状況は

$$Insu_i = \beta_2 + \beta_3 Lot_i + u_i \tag{13}$$

と予測される。一致推定量を得るために操作変数が満たすべき仮定は以下である。

仮定1. Relevance Assumption:

操作変数 *Lot* は *Insu* に因果効果がある

ここでは、くじ引きにあたったから社会保険に加入したという図式を意味する。

仮定2. Independence Assumption:

操作変数 *Lot* はランダムに割り当てられている

換言すれば、*Lot* とすべての欠落変数は独立でなければならない。

仮定3. Exclusion Restriction:

操作変数 Lot は $Insu$ を通じてのみ $Health$ に効果がある

したがって、 Lot から直接 $Health$ への効果はないことが条件となる⁶⁾。上記の設定では、操作変数 Lot は全ての仮定を満たすことが確認できる。そのうえで、

$$\begin{aligned} & \text{くじ引き当選が健康に与える効果} \\ &= (\{ \text{くじ引き当選が社会保険加入に与える効果} \} \\ & \quad \{ \text{社会保険加入が健康に与える効果} \}) \end{aligned}$$

であることを利用して式を変形すれば、操作変数を用いた一致推定量は、

$$\begin{aligned} & \text{社会保険が健康に与える効果} \\ &= \frac{\text{くじ引き当選が健康に与える効果}}{\text{くじ引き当選が社会保険加入に与える効果}} \end{aligned} \quad (14)$$

となる。操作変数法では、操作変数が説明変数に与える効果（社会保険加入が健康に与える効果）を第1段階(first stage)と呼び、操作変数が被説明変数に与える効果（くじ引き当選が健康に与える効果）を誘導型(reduced form)と呼ぶ。アウトカムを Y_i 、説明変数を D_i 、操作変数を Z_i と置き換えフォーマルに表記すれば、第1段階は $E[D_i|Z_i=1] - E[D_i|Z_i=0]$ （以下 ϕ ）、誘導型は $E[Y_i|Z_i=1] - E[Y_i|Z_i=0]$ （以下 ρ ）となり、局所平均効果は λ 、

$$\lambda = \frac{\rho}{\phi} = \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]} \quad (15)$$

となる。操作変数 Z_i と説明変数 D_i の組み合わせに着目すれば、以下の4つの部分母集団を得る。

$$Compliers : D^{Z=0} = 0 \cap D^{Z=1} = 1 \quad (16)$$

$$Defiers : D^{Z=0} = 1 \cap D^{Z=1} = 0 \quad (17)$$

$$AlwaysTakers : D^{Z=0} = 1 \cap D^{Z=1} = 1 \quad (18)$$

$$NeverTakers : D^{Z=0} = 0 \cap D^{Z=1} = 0 \quad (19)$$

Compliers とは「 Z に応答する人たち」のことであり、くじにはずれば社会保険に加入せず、くじに当たれば社会保険に加入する。*Defiers* とは「 Z とは逆の反応をする人たち」のことであり、くじにはずれば社会保険に加入し、くじに当たれば社会保険に加入しない。*AlwaysTakers* とは「常に処置を受ける人たち」のことであり、くじに当たろうがはずれようが社会保険に加入する。*NeverTakers* とは「常に処置を受けない人たち」のことであり、くじに当たろうがはずれようが社会保険に加入する。それでは操作変数による推定量はどの部分母集団への効果なのだろうか。まず、操作変数の仮定である Relevance Assumption から *AlwaysTakers* と *NeverTakers* は除外される。さらに、現実には *Defiers* が存在することはないだろうと考えられる (Angrist and Pischke 2014)。 Z と逆の動きをするのであればそもそも社会保険加入のチャンスがあるくじ引きに参加しないからだ。しかしながら、政策プログラム以外の場合にも操作変数を用いることを考えると、以上の理屈は厳密には $D^{Z=1} \geq D^{Z=0}$ という単調性(monotonicity)の仮定が必要となる⁷⁾。以上から、(15) 式の操作変数法による推定量は *Compliers* に対する局所的因果効果(Local Average Treatment Effect: LATE) であることが分かる。以下では、分析手続きについて Angrist and Pischke(2009) が推奨する確認事項を提示する。第1に、第1段階の結果を報告し、係数の大きさや符号

は予想通りかを確認する。第2 に、第1 段階の回帰式において操作変数が強い効果をもつことを確認する。定理ではないが、Stock et al.(2002) はF 値が10以上であると安全圏だとしている。第3 に、最も良い操作変数を1 つ選び丁度識別⁸⁾ の推定値を報告する。丁度識別の操作変数はメディアンの意味で不偏性があるため、弱操作変数の批判を受けにくい。第4に、制限情報最尤推定量(LIML)による過剰識別推定値を確認する。制限情報最尤推定量は2 段階最小二乗法より精度は低いもののバイアスは小さく、両推定結果が似ていれば好ましい。第5 に、誘導型の回帰式における操作変数の係数、t 値、F 値を確認する。誘導型で因果効果が確認できなければ、おそらく除外制約を成立しているだろう (Angrist and Krueger 2001; Chernozhukov and Hansen 2008)。

4. 回帰不連続デザイン

回帰不連続デザイン (Regression Discontinuity Design: RDD) とは、1990 年後半から多用されており、制度や恣意的なルール引き起こす分断点を利用して因果効果を推定する手法である。例えば医療保険制度を考えれば、日本では75 歳以上の窓口負担は後期高齢者医療制度により原則1 割 (75 歳未満は原則2、3 割) となる。この分断点についてなぜ75 歳なのかと聞かれば、政治的・歴史的・財政的等の様々な理由で決定されたのだが、RDD が着目するのは、生きている限り年齢は自動的に積み重ねていくものなので、75 歳近傍では窓口負担の変更という処置が実験のようなランダム割り当てに近いという状況である。処置がランダムに近い状況で割り当てられているのであれば、分断点近傍の人たちのアウトカムを比較すれば因果効果を推定できるというのがRDDの基本的発想である⁹⁾。RDD はSharp RDD(SRD) とFuzzy RDD(FRD) に分類できるが、これは分断線の左右において処置が決定的に決まる (SRD) か確率的に決まる (FRD) かで区別できる。後期高齢者医療制度の場合には、原則として75 歳以上は窓口負担が必ず1 割になるのでSRD である。他方、仮に75 歳以上で急激に所得が低くなる国が存在し、所得が低い場合には窓口負担が1 割となる制度があった場合には、75 歳以上→ 所得低い→ 窓口負担1 割というチャネルを強く予測するためFRD となる。つまりFRD は操作変数法を用いるのである。以下ではSRD についてその理論と分析事例を紹介する。SRD は処置 D_i が説明変数 x_i の決定関数¹⁰⁾ であり、かつ分断点 x_0 で不連続に決定するので、

$$D_i = \begin{cases} 1 & (x_i \geq x_0) \\ 0 & (x_i < x_0) \end{cases}$$

となり、 y_i は以下で表記できる。

$$y_i = \beta_0 + \beta_1 x_i + \rho D_i + \epsilon_i \quad (20)$$

x_i が n 次の場合も仮定すると、

$$y_i = \gamma + \delta_1 x_i + \delta_2 x_i^2 + \dots + \delta_n x_i^n + \rho D_i + \epsilon_i \quad (21)$$

となり、さらに D_i は x_i の決定関数であるので回帰モデルは以下となる。

$$E[Y_i | x_i] = E[Y_{0i} | x_i] + (E[Y_{1i} | x_i] - E[Y_{0i} | x_i]) D_i \quad (22)$$

上式は $\tilde{x}_i \equiv x_i - x_0$, $\delta_1^* = \delta_{11} - \delta_{01}$, $\delta_2^* = \delta_{12} - \delta_{02}$, ..., $\delta_n^* = \delta_{1n} - \delta_{0n}$ とすると、

$$y_i = \gamma + \delta_{01}\tilde{x}_i^1 + \delta_{02}\tilde{x}_i^2 + \dots + \delta_{0n}\tilde{x}_i^n + \rho D_i + \delta_1^* D_i \tilde{x}_i + \delta_2^* D_i \tilde{x}_i^2 + \dots + \delta_n^* D_i \tilde{x}_i^n + \epsilon_i \quad (23)$$

と書き直せる。これまでにみた共変量を統制する手法とは異なり、処置が継続変数の決定関数であるため、RDD は分断点付近では共変量とは無関係に因果効果を推定できるのである。以下では、Lee et al.(2004) のデータ¹¹⁾を用いて RDD の基礎的方法を確認する。Lee et al.(2004) は選挙において、a) 有権者が候補者の政策選好の影響を与えるのか、b) それとも単に有権者は政党間で異なる政策を選んでいるのかに着目する。具体的には、民主党議員が当選したならば、その議員は議会においてよりリベラルな法案に賛成しているだろうという問いを検証している。アメリカでは各議員が法案に賛成したか否かが記録されている点呼投票 (roll call vote) のデータが存在し、その記録を用いて 0 から 100 のリベラル度合いを示すインデックス (Americans for Democratic Action Score: ADA Score) が作成されている。従って、 t 期の民主党の得票率 x_i が 50% (x_0) を超えるならば、 $t+1$ 期の ADA スコアはどの程度上昇するかを検証する。図1 は民主党得票率を5次の項まで投入したモデルで近似した曲線を引いたものである。

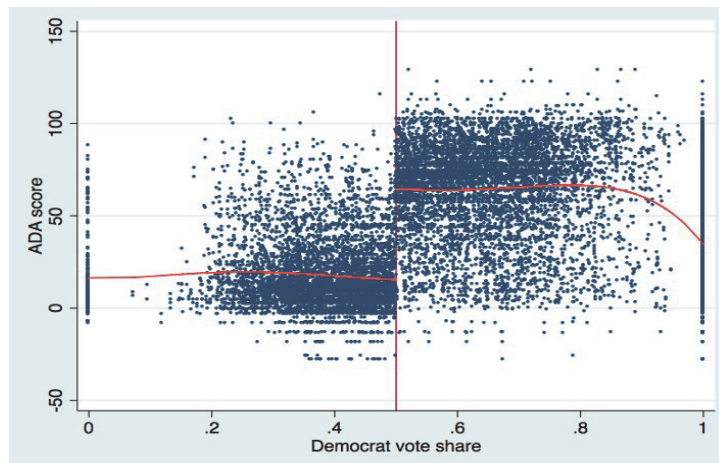


図1 t 期の民主党得票率と $t+1$ 期の ADA スコア

注) Lee et al.(2004) を基に筆者作成

得票率50% の分断点で ADA スコアがジャンプしているのが確認できるが、これが因果効果 ρ である。RDD で必要となる作業は図1 のように分断点においてジャンプがあるかどうかを確認することであり、ここでジャンプが確認されなければ因果効果はないだろう。次に x_i を $x_0 = 50$ で中心化した式(23)を OLS で推定すると $\rho = 47.43 (SE = 2.04)$ であり分断点におけるジャンプは有意に正であることが分かる。以上が RDD の概略であるが、実際に分析する際には以下に注意する必要がある。

(イ) 継続変数の操作可能性がないことを確認する

処置が分断点付近においてあたかもランダムのように生じていることを担保するのは、継続変数の操作

可能性がないことである。民主党にとって得票率を49%にするか51%にするかを操作するのは不可能である。実践的には継続変数のヒストグラムやカーネル密度をみることで、分断点近傍で分布が大きく変化しないかを確認する。よりフォーマルには、McCrary(2008)が提示したDensity Testによって検定が可能である。

(ロ) 分断点近傍の2群で共変量に大きな変化がないかを確認する

操作可能性のない継続変数を分断点におけるランダム割り当てとみなすためには、分断点近傍の処置群($D_i = 1$)と対照群($D_i = 0$)で共変量に違いがあつてはならない。この点は、多くの場合にグラフを書くことで確認できる。すなわち、横軸に継続変数を、縦軸に共変量をとった散布図を描き、分断点近傍でジャンプがないかを確認することが不可欠である。

(ハ) 分断点近傍のバンド幅に注意する

分断点近傍のバンド幅をどの程度に設定するかによって分析結果は変わってくる。パラメトリックな推定では、様々な高次多項式を投入してモデル比較を行うのが一般的である(Lee and Lemieux 2010)。ここで結果が大きく変わるようであれば問題だが、近年ではGelman and Imbens(2014)が高次多項式の投入は避けることを推奨している。ノンパラメトリックな推定では、パラメトリックな推定におけるバンド幅より狭くして平均値の比較を行うが、これはサンプルサイズが十分に確保されていなければならない。よりフォーマルな方法としては、Imbens and Kalyanaraman(2012)が適切なバンド幅の選び方を定式している。

以上に概観したRDDは、自然実験を利用した内的妥当性の高い推定結果となる。その一方で、RDDは分断点近傍(subpopulation)の局所平均効果になっているため、分断点近傍から離れた場合に同様の効果があるかは必ずしも自明ではない。すなわち、特定のバンド幅以外にも妥当する結果であるのかという外敵妥当性については検討の余地があるということだ。近年では、Hainmueller et al.(2015)やAngrist and Rokkanen(2016)のように、RDDの外敵妥当性について検討する研究も出てきている。

5. 差分の差法

差分の差法とも訳されるDifferences-in-Differences(以下DiD)は、パネルデータや繰り返しクロスセクションデータを用いて広く採用されている手法である。介護保険制度の導入が健康にどのような影響を与えたのかについて因果効果を知りたいとしよう。介護保険導入前後の2期間($t = 0, 1$)、と介護保険サービスを受けたか否か($D = 0, 1$)を想定し潜在的アウトカム y_{Dt} を設定すると、介護保険導入前後での因果効果 δ_{DD} は、

$$\delta_{DD} = (y_{11} - y_{10}) - (y_{01} - y_{00}) \quad (24)$$

$$= (y_{11} - y_{01}) - (y_{10} - y_{00}) \quad (25)$$

となる。式(24)は「処置群での介護保険導入前後の差」と「対照群での介護保険導入前後の差」の差分であり、式(25)は「介護保険導入後の処置群と対照群の差」と「介護保険導入前の処置群と対照群の差」

である。処置群ダミーを D 、介護保険導入後ダミーを P とすると、 δ_{DD} は図2における $E[Y_1(1)Y_0(1)|D=1]$ に該当し、さらに一般的には以下の回帰式で推定される。

$$Y_{dt} = \alpha + \beta D + \gamma P + \delta_{DD} (D * P) + \epsilon \quad (26)$$

なぜ処置群ダミーと介護保険導入後ダミーの交差項の係数が δ_{DD} となるかは、表1から明らかである。

DiD 推定量を式(26)から導くためには、ある仮定を満たしていなければならない。その仮定とは、処置の期間を除いては、処置群と対照群には共通のトレンド (common trend) が存在していることだ。

表 1 DiD 推定

	t_1	t_0	$t_1 - t_0$
D_1	$\alpha + \beta + \gamma + \delta_{DD}$	$\alpha + \beta$	$\gamma + \delta_{DD}$
D_0	$\alpha + \gamma$	α	γ
$D_1 - D_0$	$\beta + \delta_{DD}$	β	δ_{DD}

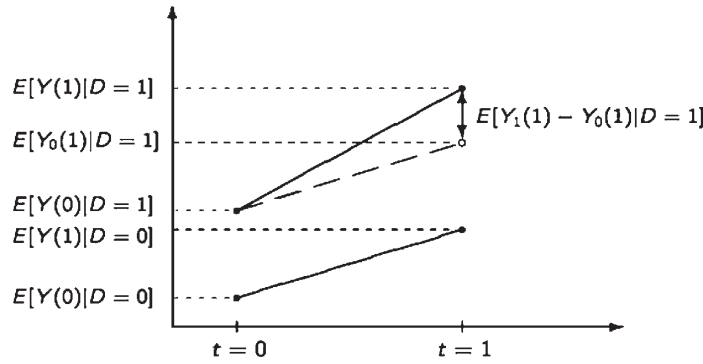


図 2 潜在的アウトカム

すなわち、横軸に時間、縦軸にアウトカムをとった場合に、処置の期間を除けば処置群と対照群のトレンド（傾き）は同じになっていることが必要である。なぜこの仮定が必要となるかは以下の通りである。まず式(25)を期待値で書き直すと、

$$\begin{aligned} & \{E[Y(1)|D=1] - E[Y(1)|D=0]\} - \{E[Y(0)|D=1] - E[Y(0)|D=0]\} \\ &= \{E[Y(1)_1|D=1] - E[Y(1)_0|D=0]\} - \{E[Y(0)_0|D=1] - E[Y(0)_0|D=0]\} \end{aligned} \quad (27)$$

共通トレンドの仮定とは、

$$\begin{aligned} E[Y_0(1) - Y_0(0)|D=1] &= E[Y_0(1) - Y_0(0)|D=0] \\ E[Y_0(1)|D=0] &= E[Y_0(1)|D=1] - E[Y_0(0)|D=1] + E[Y_0(0)|D=0] \end{aligned} \quad (28)$$

なので、式(28)を式(27)に代入して、

$$\begin{aligned} & \{E[Y(1)_1|D=1] - (E[Y_0(1)|D=1] - E[Y_0(0)|D=1] + E[Y_0(0)|D=0])\} \\ & \quad - \{E[Y_0(0)|D=1] + E[Y_0(0)|D=0]\} \\ &= E[Y_1(1) - Y_0(1)|D=1] \end{aligned} \quad (29)$$

となり、共通トレンドの仮定があることで処置群における平均因果効果を得ることができる。実践的に

はグラフを作成し共通トレンドの仮定をチェックすることが多いが、現実には共通トレンドを満たさないこともあり、その場合には傾向スコアマッチング等による補正が応用される¹²⁾。DiD を用いた因果効果の推定にあたっては、共通トレンド仮定の確認は必須作業である。

6. 因果推論をめぐる議論

これまで、因果推論の具体的方法であるランダム化比較実験、回帰分析、操作変数、RDD、DiD を概観した。社会科学の査読付きジャーナルでは、こうした手法を用いた分析が蓄積しており、社会学でも因果推論を特集する論文や書籍も増加傾向にある (Sobel 1998; Winship and Morgan 1999; Goldthorpe 2001; Gangl 2010; Bollen 2012; Yamaguchi 2012; Morgan and Winship 2014)。他方、因果推論に対しては以下のような批判や議論がある。

I. 因果推論では因果効果ばかりに着目している

因果推論では、着目する説明変数がアウトカムに与える因果効果を明らかにする研究が多い (Imbens 2010; Gelman and Imbens 2013)。これに対して、因果効果を解明する分析 (cause of effect) ばかりではなく因果効果が存在するメカニズムの解明 (effect of cause) に比重を置くべきだとする批判がある (Heckman 2005; Deaton 2010)。社会学者である Goldthorpe (2001) も、社会学の使命は「因果メカニズムの解明である」と述べている。一昔前は、因果効果の分析が多いという理由で因果推論が敬遠されていたかもしれないが、近年では因果メカニズムを明らかにする手法が急速に発達している (Imai et al. 2011; VanderWeele 2015)。因果メカニズムの解明は紛れもなく因果推論の守備範囲なのだ¹³⁾。

II. 因果推論では個体の異質性が考慮されていない

因果推論では平均因果効果を推定するものが多く、その効果が個体間で均質であることに批判がよせられてきた (Imai and Strauss 2011)。しかしながら近年では、因果効果の異質性 (Heterogenous Treatment Effect) に着目する研究が多く (Morgan 2001; Angrist 2004; Brand and Xie 2010; Imai and Strauss 2011; Breen et al. 2015)、分析手法にもかなりのヴァリエーションがある。因果推論だから個体間の異質性が考慮できないというのは明白な誤りである。

III. 因果推論で設ける仮定の妥当性があやしい

因果推論では分析過程で様々な仮定が設けられるが、その仮定の妥当性をめぐりしばしば議論がなされる。代表的なものとしては、Stable Unit Treatment Value Assumption (SUTVA) がある。SUTVA は処置 D の y_i に対する効果は他人 j の処置状況に依存しないことを意味する¹⁴⁾。この仮定は個人や組織の相互作用を考察してきた社会学には大きな制限となりうる。現段階では、SUTVA 違反がある場合には、分析単位をよりマクロベルにすることや、因果推論において通常なされる局所平均効果の仮定を緩めることが提案されている (Sobel 2006; Heckman et al. 2006; Morgan and Winship 2014)。

以上の議論は因果推論そのものに対する批判というよりも、因果推論の課題を挙げていると理解すべきである。近年の因果推論の発展は、上記の課題を克服しようと目指しいほどの研究業績が蓄積されている。もちろん、全ての分野で因果推論が主流になるべきだと説いているわけではなく、分野によってはモデルフィット型の分析が求められることもあるだろう。しかしながら、近年の因果推論の著しい分析手法の発展を考慮すれば、様々な応用可能性があり、データ分析をするものにとって因果推論は避けて通れないツールとなるだろう。翻って、本稿で挙げたような基礎的な分析手法が共有化されることによって、社会学の研究領域が経済学や政治学に浸食され過ぎるのを防ぐとも思われる。

参考文献

- [1] Angrist, J. and A. Krueger. 1999. "Empirical Strategies in Labor Economics." O. Ashenfelter and D. Card. eds., *Handbook of Labor Economics Vol.3* Amsterdam: North Holland, 1277-1366.
- [2] Angrist, J. D., and A. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15(4): 69-85.
- [3] Angrist, J. 2004. "Treatment Effect Heterogeneity in Theory and Practice." *The Economic Journal* 114: 52-83.
- [4] Angrist, J. and J. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, Princeton University Press.
- [5] Angrist, J. and J. Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton, Princeton University Press.
- [6] Angrist, J. and Rokkanen, M. 2016. "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff". *Journal of the American Statistical Association* forthcoming.
- [7] Bollen, K. 2012. "Instrumental Variables in Sociology and the Social Sciences." *Annual Review of Sociology* 38: 37-72.
- [8] Brand, J. and Y. Xie. 2010. "Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education." *American Sociological Review* 75: 273-302.
- [9] Breen, R., Choi, S., and A. Holm. 2015. "Heterogeneous Causal Effects and Sample Selection Bias." *Sociological Science* 2: 351-369.
- [10] Chernozhukov, V. and C. Hansen. 2008. "The Reduced Form: A Simple Approach to Inference with Weak Instruments." *Economic Letters* 100(1): 68-71.
- [11] Deaton, A. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48(2): 424-455.
- [12] Duncan, O. 1984. *Notes on Social Measurement: Historical and Critical*. New York: Russell Sage Foundation.
- [13] Gangl, M. 2010. "Causal Inference in Sociological Research." *Annual Review of Sociology* 36: 21-47.
- [14] Gelman, A. and G. Imbens. 2013. "Why ask Why? Forward Causal Inference and Reverse Causal Questions" *NBER Working Paper* No. 19614.
- [15] Gelman, A. and Imbens, G. 2014. "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs." *NBER Working Paper* No. 20405.
- [16] Goldthorpe, J. 2001. "Causation, Statistics, and Sociology." *European Sociological Review* 17: 1-20.
- [17] Hainmueller, J., Hall, A. and J. Snyder. 2015. "Assessing the External Validity of Election RD Estimates: An Investigation of the Incumbency Advantage." *Journal of Politics* 77(3): 707-720.
- [18] Heckman, J. 2005. "The Scientific Model of Causality." *Sociological Methodology* 35: 1-97.
- [19] Heckman, J., S. Urzua, and E. Vytlačil. 2006. "Understanding Instrumental Variables in Models with Essential

- Heterogeneity.” *Review of Economics and Statistics* 88: 389-432.
- [20] Imai, K., Keele, L., Tingley, D., and T. Yamamoto. 2011. “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies.” *American Political Science Review* 105(4): 765-789.
- [21] Imai, K. and A. Strauss. 2011. “Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-out-the-vote Campaign.” *Political Analysis* 19: 1-19.
- [22] Imbens, G. 2010. “Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009).” *Journal of Economic Literature* 48(2): 399-423.
- [23] Imbens, G., and Kalyanaraman, K. 2012. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *The Review of Economic Studies* 79(3): 933-959.
- [24] Lee, D. S., Moretti, E., and Butler, M. J. 2004. “Do Voters Affect or Elect Policies? Evidence from the U. S. House.” *The Quarterly Journal of Economics* 119(3): 807-859.
- [25] Lee, D. S., and Lemieux, T. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48(2): 281-355.
- [26] McCrary, J. 2008. “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test” *Journal of Econometrics* 142(2): 698-714.
- [27] Morgan, S. 2001. “Counterfactuals, Causal Effect Heterogeneity, and the Catholic School Effect on Learning.” *Sociology of Education* 74: 341-74.
- [28] Morgan, S. and C. Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- [29] Rubin, D. 1991. “Practical Implications of Modes of Statistical Inference for Causal Effects and the Critical Role of the Assignment Mechanism.” *Biometrics* 47(4): 1213-1234.
- [30] Sobel, M. 1998. “Causal Inference in Statistical Models of the Process of Socioeconomic Achievement: A Case Study.” *Sociological Methods and Research* 27: 318-48.
- [31] Sobel, M. 2006. “What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Inference.” *Journal of the American Statistical Association* 101: 1398-407.
- [32] Stock, J., Wright, J. and M. Yogo. 2002. “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments.” *Journal of Business and Economic Statistics* 20(4): 518-529.
- [33] Stouffer, S. 1950. “Some Observations on Study Design.” *American Journal of Sociology* 55: 355-61.
- [34] VanderWeele, T. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford: Oxford University Press.
- [35] Winship, C. and S. Morgan. 1999. “The Estimation of Causal Effects from Observational Data.” *Annual Review of Sociology* 25: 659-706.
- [36] Xie, Y. 2007. “Otis Dudley Duncan's Legacy: The Demographic Approach to Quantitative Reasoning in Social Science.” *Research in Social Stratification and Mobility* 25: 141-56.
- [37] Yamaguchi, K. 2012. “Log-linear Causal Analysis of Cross-classified Categorical Data.” *Sociological Methodology* 42: 257-285.

注

- 1) 無論、計量分析の目的は因果関係の特定だけではないため、分析目的が記述的あるいは探索的な場合にはモデルフィットに重点をおいた分析手法を採用すべきであり、こうした場合に先の批判は該当しない。
- 2) 因果状態とは、実際には実現しなかった潜在変数 (potential outcome) と実際に実現した変数 (observed outcome) の関係を指す。

- 3) 性別や人種が結果に与える理想的な実験は一見不可能のように思えるが、性別や人種をめぐる差別は、差別する者が相手を「特定の性別や人種である」という認識から生じる。この点をうまく利用すれば、虚偽の履歴書や個人情報等を郵送し雇用者の反応を探るといった研究もデザインでき、実際にこうした研究はフィールド実験として存在している (Bertrand and Mullainathan 2004; Fryer and Levitt 2004; Western 2009)。
- 4) 介護保険サービスを受ける人は受けない人に比べて、そもそも身体機能が低下している人が多いと考えられ、その場合には $E[Y_{0i}|D_i = 1] \neq E[Y_{0i}|D_i = 0]$ となる。
- 5) $Cov[PInsu_i, Health_i] = 0$ から $\beta_1 = \frac{Cov[Health_i, PInsu_i]}{Var[PInsu_i]}$ となり、式 (3) を上式に代入すれば $E[e_i|PInsu_i] = E[E[e_i|Gene_i, PInsu_i]|PInsu_i] = 0$ が成立し、 $Cov[PInsu_i, e_i] = 0$ なので、 $\beta_1 = \beta_2 + \beta_3 \frac{Cov[PInsu_i, Gene_i]}{Var[PInsu_i]}$ となる。式 (2) と式 (3) の β_1 と β_2 は異なるパラメータとして識別され、 β_1 の推定量は上式の右辺に確率収束するため β_2 の一致推定量とならないことが証明される。
- 6) Exclusion Restriction (除外制約) をフォーマルにテストする方法は存在しないが、誘導系の回帰において操作変数があり意図しないことを確認することは除外制約の成立を示唆するため、実践的には利用されている。
- 7) 換言すれば、操作変数が説明変数に与える効果は一定の方向 (正か負) であるということである。
- 8) 操作変数が 1 つの場合を丁度識別、操作変数が説明変数より多い場合を過剰識別という。
- 9) 実際に窓口負担が 1 割になることを利用した分析が Shigeoka(2014) である。
- 10) RD では着目する説明変数のことを強制変数 (forcing variable) や継続変数 (running variable) と呼ぶため、以下ではこの用語を使用する。
- 11) データは著者のひとりである Enrico Moretti の HP(<http://eml.berkeley.edu/~moretti/data3.html>) から入手できる。
- 12) 共通トレンドの仮定を一般化もしくは修正した代表的なものとしては Athey and Imbens(2006) や Abadie(2006) があげられる。またモンテカルロシミュレーションを用いて DiD 推定がどの程度信頼できるかを検討したものとしては Bertland et al.(2004) がある。
- 13) 反実仮定の枠組みによる因果推論では、大きく潜在結果変数モデル (Potential Outcome Model) と有向グラフアプローチ (Directed Graph Approach) に分類できる (Morgan and Winship 2014)。前者はルービン (D. Rubin)、後者はパール (J. Pearl) によって牽引されてきたが、本稿では POM の枠組みに依拠している。
- 14) 具体的には、予防接種の予防効果は単に自らが予防接種をしたときよりも、周囲のできるだけ多くの人が予防接種を受けていたほうが効果があるかもしれない。このような場合には SUTVA は成立しない。

Causal Inference in Theory and Practice

Shoki OKUBO

Abstract:

This paper introduces theory and practice of causal inference and discusses how to estimate causal effects. We first discuss how randomized experiments allows us to assess causal effects and then turn to observational studies. We lay out the assumptions needed for causal inference and describe the leading analysis methods, including regression, instrumental variables, regression discontinuity and differences-in-differences.

Key Words : Causal Inference, Randomized Controlled Trial, Instrumental Variables, Regression Discontinuity Design, Differences-in-Differences